



Intensity-based Similarity Measures Evaluation for CT to Ultrasound 2D Registration

Zulma Sandoval, Jean-Louis Dillenseger

► To cite this version:

Zulma Sandoval, Jean-Louis Dillenseger. Intensity-based Similarity Measures Evaluation for CT to Ultrasound 2D Registration. *Innovation and Research in BioMedical engineering*, 2013, 34 (4-5), pp.278-282. 10.1016/j.irbm.2013.08.001 . inserm-00879579

HAL Id: inserm-00879579

<https://www.hal.inserm.fr/inserm-00879579>

Submitted on 4 Nov 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Intensity-based Similarity Measures Evaluation for CT to Ultrasound 2D Registration

Sandoval Zulma^a, Dillenseger Jean-Louis^a,

^a*Inserm, U1099, Rennes, F-35000, France ; Université de Rennes 1, LTSI, Rennes, F-35000, France*

Abstract

Ultrasound is a non-invasive image modality which allows for real time acquisition. Nevertheless, the low quality of the acquired images makes this a difficult-to-interpret modality during surgical procedures. To overcome this, the registration of ultrasound images with pre-operative CT or MR images has been routinely used to fuse complementary information. This work presents the evaluation of eight similarity measures used in the registration of ultrasound and CT images of the left atrium and the pulmonary veins. Each intensity-based similarity measure was evaluated computing its accuracy, capture range, distinctiveness of the optimum, risk and non-convergence and number of minima. The results show, that the Woods criterion presents a globally better performance than the other similarity measures. This is especially true for the accuracy and distinctiveness of the optimum indicators. Preprocessing US images does not improve the performance of all similarity measures, except for Woods criterion that shows the optimal accuracy.

Keywords: Similarity measures, rigid registration, ultrasound registration, multimodality registration, evaluation

1. Introduction

Atrial fibrillation is a cardiac arrhythmia caused by abnormal electrical discharges in the atrium. Ablation procedures have proven to be some of the most effective methods in treating atrial fibrillation [1]. These procedures aim to establish a line of lesions around the pulmonary veins in order to block trigger points of atrial fibrillation. Ultrasound-guided HIFU is a minimally-invasive alternative to other ablation techniques[2]. In this procedure, the

ablation path is defined in a preoperative stage using 3D cine CT scans. The ablation itself is performed under 2D ultrasound (US) guidance. The 2D-3D registration of intra-operative US images to the pre-operative CT images is then needed in order to transfer and follow the ablation path in the pre-operative context.

Beside feature-based registration methods which need a first image segmentation step, an intensity-based cardiac US to cardiac CT image registration method has been described in [3]. The authors of this paper proposed a spatio-temporal registration of the beating heart using mutual information (MI) metric to drive the spatial alignment. This technique has been improved by performing an intensity-based registration on only the relevant features of the US image after median filtering and low intensity thresholding [4]. The question of MI validity thus arises. In CT images, tissues are characterized by intensity distribution. However, in US images, each tissue is inherently characterized by a specific spatial distribution of speckle rather than a specific distribution of gray levels. Therefore, the use of an intensity-based SM in a multimodality image registration problem involving US images is still an open question. Beside MI, other intensity-based SMs have also been proposed in the literature. Some of these metrics could be more adequate than MI in driving the registration.

In this paper, we propose to validate and quantify the usability and impact of such SMs for US to CT registration. The evaluation of five properties of eight cited intensity-based SMs between US and CT images is performed. This evaluation aims to find the best SM or optimal combination of SMs to be included in our registration framework. The ability to improve the behavior of SMs using the pre-processing procedure proposed in [4] is also evaluated. The protocol used in this evaluation requires a gold standard registration. We thus propose to use synthetic CT and US images with a known gold standard.

This document is structured as follows: in section 2, the synthetic images, similarity measures and protocol used in the evaluation are presented. In section 3, the experimental results are shown and discussed.

2. Materials and Methods

2.1. Images

Because the evaluation requires a registration gold standard, we choose to use realistic synthetic data. Two sets of synthetic images were built for

use in the evaluation: 1) a geometrical form composed of an oval on a plain background [Fig. A.1-(a) and (b)]; and 2) a set of four synthetic phantom images of the left atrium and pulmonary veins (LAPVs), more related to our medical application. These four phantom images are composed of two images with different fields of view (FOV) from a male cryosection of the human thorax and two images with different FOV from a female cryosection of the human thorax. The cryosections were taken from the Visible Human Project ¹. As an example, Fig. A.1-(c) shows a slice from the male dataset. In this section LAPVs images are segmented and labeled into six types of tissues: air, fat, bone, muscle, water and blood (A.1-(d)). Synthetic CT images were obtained by applying the Hounsfield value corresponding to the type of tissue modulated by Gaussian noise (Fig. A.1-(e)). In simulating US images, each tissue is first characterized by an acoustical impedance and a spatial distribution of speckle. This information is set as an input to the US image simulator [5] (Fig. A.1-(f)). More details can be found in the appendix.

2.2. Registration procedure and similarity measures

The registration procedure of two images of the same scene acquired at different times, from different points of view and/or by different sensors, aims to align them into a common referential. Usually, one image is fixed as a reference and the other is moved and compared with it. The registration procedure consists of 1) applying a geometric transformation to the moving image, 2) comparing similarities between the two images by using a SM and 3) optimizing the geometric transformation in order to maximize similarity between the two images. Ideally, the SM has an optimum at the set of parameters that best align the moving image, and values that decrease monotonically with the distance to the optimum. In practice, local minima may be present close to the optimum. So the good choice of a SM adapted to our medical application is crucial.

In this paper, eight intensity-based SMs are evaluated. Six of these SMs use the information from the histogram of images while two of them use the spatial information and intensity values. They are: mutual information (MI) [6], normalized mutual information (NMI) [7], entropy correlation coefficient (ECC) [6], joint entropy (H) [6], point similarity measure based on MI (PSMI)

¹http://www.nlm.nih.gov/research/visible/visible_gallery.html

[8], histogram energy (E) [9], correlation ratio (CoR) [10] and Woods criterion (WC)[11].

2.3. Evaluation protocol

The protocol consists of a statistical and systematic evaluation of the SM behavior around the registration gold standard. This protocol proposed by Skerl [12] carries out a thorough performance evaluation of the SMs independently of the optimization method used to find the optimal alignment. The protocol needs to know the gold standard i.e., the underlying transformation between the fixed and moving images. Three steps are then needed to perform an evaluation:

1. The definition of statistical significant sample points within the parametric space around the gold standard. The parametric space is defined according to the dimensions, normalization values and amount of points used in the computation of the SM. The dimension is defined by the amount of parameters used in the transformation (two translations and one rotation angle, in the context of 2D rigid transformation). The parametric space is normalized in such a way that a change along each axis, produces almost the same impact on the transformation magnitude. After normalization, Euclidean metrics may be used to measure the distance between the values of the SM. The protocol generates a number (M) of sampling points equally distributed on (N) lines randomly localized. All these lines intersect at the origin of the parametric space, the gold standard. Sampling points lie inside of a hypersphere of radius R in the parametric space.
2. SM values around the gold standard are computed using geometrical parameters corresponding to specific sample points, as defined at Step 1.
3. SM behavior analysis around the gold standard. For this, five properties are measured from the SM values obtained for each line during the second step (Fig. A.2). The accuracy (ACC) of the registration corresponds to the root-mean-square of the difference between the position of the optimal value (P_{MAX}) and the gold standard (P_{GS}), for all lines (1).

$$ACC = \sqrt{\frac{1}{N} \sum_{n=1}^N \|P_{n,max} - P_{GS}\|^2} \quad (1)$$

The distinctiveness of the optimum (DO) measures the visibility of the maximum value against the values around it. It is calculated as the average of the difference between twice the maximum value and the sum of the two values symmetrically located at their two sides at a given distance, as is defined in (2).

$$DO(r) = \frac{1}{2rN} \sum_{n=1}^N 2SM(X_{n,max}) - SM(X_{n,max-k}) - SM(X_{n,max+k}), \quad (2)$$

with $r = k * \delta$, δ the distance between two sampling points in a line ($\delta = 2R/M$), and k is the number of steps ($k = 1, 2, \dots$). The number of minima (NOM) is the amount of local minima of the SM around the gold standard. The capture range (CR) of the similarity measure is the distance from the nearest local minima (P_{NLM}) to the maximum value of the SM (P_{MAX}). The risk of non-convergence (RON) of the registration is the average of the positive differences between the SMs at consecutive sampling points, as is defined in (3).

$$RON(r) = \frac{1}{2rN} \sum_{n=1}^N \sum_{P_{MAX}-k}^{P_{MAX}+k} d_{n,m}, \quad (3)$$

$$d_{n,m} = \begin{cases} SM(X_{n,m-1}) - SM(X_{n,m}), & \text{if } m < P_{MAX} \text{ and } SM(X_{n,m-1}) > SM(X_{n,m}) \\ SM(X_{n,m+1}) - SM(X_{n,m}), & \text{if } m > P_{MAX} \text{ and } SM(X_{n,m+1}) > SM(X_{n,m}) \\ 0, & \text{if } otherwise. \end{cases}$$

These five properties are good indicators in predicting SM behavior for a good registration (when ACC, RON and NOM values are as low as possible and CR and DO values as high as possible).

3. Results

The evaluation was carried out on the Oval and four LAPVs images. In each of these cases, we performed the registration of the CT image with the corresponding raw US images and also with the US images preprocessed as in [4]. The values used to define sampling points were set as is suggested in ²: $N=7$, $M=100$ and $R=0.3$. Using these values, the distance between two

²<http://lit.fe.uni-lj.si/contents/tools/SimilarityEvaluation/>

consecutive points along a line was much smaller than the smallest pixel size. Table A.3 shows image sizes, normalization values, and distances between two consecutive points along a line, for each of images used in the test. The unit for rotation was chosen in such a way that a rotation around one corner of the image causes the mean shift of all pixels equal to a translation unit. For rotation and translation, we used a rigid transformation and a linear interpolation method. We used 256 bins in the computation of different SMs.

The results of the Oval will be noted as Ov (US raw data) and Ov1 (preprocessed US data) in the rest of the paper. The LAPVs images had relatively similar results. We then chose to present the mean values of these LAPV images measures and noted them as L (US raw data) and L1 (preprocessed US data). Tables A.1 and A.2 give the ACC, RON, NOM, CR and DO indicator properties of the eight SMs for the Oval and LAPV images respectively.

It can be observed in Tables A.1 and A.2 that WC presents a globally better performance than the other SMs, for both LAPV and Oval images. This is especially true for the ACC and DO indicators. MI and CoR equally give globally high performances, regardless of image content. NMI has a high ACC and the smallest NOM for Oval. Results show that MI can be used in the registration of CT to US images as has been done in [4]. However, results also show that preprocessing the images does not improve the performance. If we compare the behavior between the raw and preprocessed images, it can be concluded that the performance is almost the same for all SMs except for WC. WC has a optimal ACC for both Oval and LAPV preprocessed images. All SMs have a small CR, as was expected for these modalities. This result is consistent with that reported in [12].

4. Conclusion

ACC, RON, NOM, CR and DO properties were evaluated on eight SMs based on gray level, for the registration of CT images with US images. It can be concluded that WC has the best performance, especially concerning Oval images. MI, NMI and CoR also have a globally high performance for both Oval and LAPV. Overall, the performance does not improve for preprocessed images. However, WC has an optimal ACC using preprocessed US images, and seems to be the most optimal SM in the case of US to CT rigid registration. Because our evaluation was independent to the optimization

process, we believe that this behavior would still be true in the case of elastic registration.

This work is part of the French CardioUSgHIFU project supported by an ANR-2011-TecSan-004.

Appendix A. Image simulation parameters

The LAPVs images were first segmented and labeled into six different tissue classes: air, fat, bone, muscle (including myocardium), water and blood with contrast medium.

The CT images were simulated by assigning the corresponding Hounsfield value to each tissue (see Table A.4). We then applied an additive Gaussian noise with a standard deviation of 400 to each value.

The US images were obtained using the simulator described in [5]. In this simulator, each tissue is characterized by its acoustical impedance (see Table A.4) and a specific spatial distribution of speckle. In this speckle model, the inter-scatterer distances are independent and randomly distributed from a gamma distribution tuned by two parameters: \bar{d} represents the mean inter-scatterer distances and α a regularity parameter (see Table A.4).

References

- [1] J. Cox, Atrial fibrillation II: Rationale for surgical treatment, *J. Thoracic Cardiovasc. Surg.*, Vol. 126, pp. 1693—1699, 2003.
- [2] S. Pichardo, K. Hynynen, New design for an endoesophageal sector-based array for the treatment of atrial fibrillation: A Parametric Simulation Study *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*, vol. 56(3) 3, pp. 600–612, 2009.
- [3] X. Huang, et.al., Dynamic 2D Ultrasound and 3D CT Image Registration of the Beating Heart, *IEEE Trans. Med. Imag.* Vol. 28, pp. 1179—1189, 2009.

- [4] X. Huang, et.al., Rapid registration of multimodal images using a reduced number of voxels, *Proc. SPIE Med. Imag.* Vol. 6141, pp. 614116-1-10, 2006.
- [5] J-L. Dillenseger, S. Laguitton, et.al., Fast simulation of ultrasound images from a CT volume, *Comp. Bio. Med.* Vol. 39, pp. 180—186, 2009.
- [6] F. Maes, et.al., Multimodality Image Registration by Maximization of Mutual Information, *IEEE Trans. Med. Imag.* Vol. 16, pp. 187—198, 1997.
- [7] C. Studholme, D. Hill, et.al., An overlap invariant entropy measure of 3D medical image alignment, *Patt. Recogn.* Vol. 32, pp. 71—86, 1999.
- [8] P. Rogelj, et.al., Point similarity measures for non-rigid registration of multi-modal data, *Comp. Vis. Imag. Underst.* Vol. 92, pp. 112—140, 2003.
- [9] M.Bro-Nielsen, Rigid registration of CT, MR and crytosection images using a GLCM framework, *Proc. CVRMed 97* Vol.1205, pp. 171—180, 1997.
- [10] A. Roche, G. Malandain, et.al., The correlation ratio as a new similarity measure for multimodal image registration, *Proc. MICCAI'98* Vol. 1496, pp. 1115—1124, 1998.
- [11] R. Woods, et.al., MRI-PET registration with automated algorithm, *J Comput. Assist. Tomogr.* Vol.17, pp. 536—546, 1993.
- [12] D. Skerl, et.al., A Protocol for Evaluation of Similarity Measures for Rigid Registration, *IEEE Trans. Med. Imag.* Vol. 25, pp. 779—791, 2006.

List of Figures

A.1	Images used for evaluation. Oval geometrical form: simulated CT (a) and US (b) images. Left atrium and pulmonary veins phantom images: cryosection of male human thorax (c); segmented and labeled into six different types of tissues: air, fat, bone, muscle(including myocardium), water and blood (d); corresponding simulated CT (e) and US (f) images for a US beam depth = 7 cm.	12
A.2	Profile of a SM along one line with M sampling points, gold standard at P_{GS} , maximum value at P_{MAX} and nearest local minimum at P_{NLM}	13

List of Tables

A.1	Value of the ACC (mm), RON ($10^7/mm$), NOM, CR(mm) and DO ($1/mm$) properties for the eight similarity measures evaluated on the raw Oval image (Ov) and the preprocessed Oval image (Ov1). Sampling step along a line = 0.0588 mm ($N=7$, $M=100$, $R=0.3$). Best values are bolded.	10
A.2	Value of the ACC(mm), RON ($10^6/mm$), NOM, CR(mm) and DO ($1/mm$) properties for the eight similarity measures evaluated on: the raw LAPV image (L) and the preprocessed LAPV image (L1). Sampling step along a line: 0.0444 mm for female images and 0.0468 mm for male images ($N=7$, $M=100$, $R=0.3$). Best values are bolded.	10
A.3	Moving image sizes, field of view (FOV), translation and rotation units of normalized parametric space and distance between two consecutive points along a line, for the five images ($N = 7$ lines, $M = 100$ sampling points and $R = 0.3\text{ mm}$). . .	11
A.4	Values of parameters used in the simulation of CT and US images.	11

Table A.1: Value of the ACC (mm), RON ($10^7/mm$), NOM, CR(mm) and DO ($1/mm$) properties for the eight similarity measures evaluated on the raw Oval image (Ov) and the preprocessed Oval image (Ov1). Sampling step along a line = 0.0588 mm (N=7, M=100, R=0.3). Best values are bolded.

	ACC		RON		NOM		CR		DO	
	Ov	Ov1	Ov	Ov1	Ov	Ov1	Ov	Ov1	Ov	Ov1
MI	0.14	0.14	2.8	2.8	177	177	0.059	0.059	14	15
NMI	0.14	0.16	2.7	2.8	20	21	0.059	0.059	11	11
ECC	0.17	0.17	2.8	2.8	100	100	0.059	0.059	11	11
H	0.21	0.21	3.1	3.1	38	38	0.059	0.059	0.87	0.85
PSMI	0.21	0.21	3.1	3.1	41	40	0.059	0.059	-0.74	-0.87
E	0.21	0.2	3.1	3.1	39	40	0.059	0.059	1.0	1.3
CoR	0.16	0.16	2.7	2.7	68	75	0.059	0.059	11	12
WC	0.19	0	2.6	0.79	77	100	0.059	0.118	15	60

Table A.2: Value of the ACC(mm), RON ($10^6/mm$), NOM, CR(mm) and DO ($1/mm$) properties for the eight similarity measures evaluated on: the raw LAPV image (L) and the preprocessed LAPV image (L1). Sampling step along a line: 0.0444 mm for female images and 0.0468 mm for male images (N=7, M=100, R=0.3). Best values are bolded.

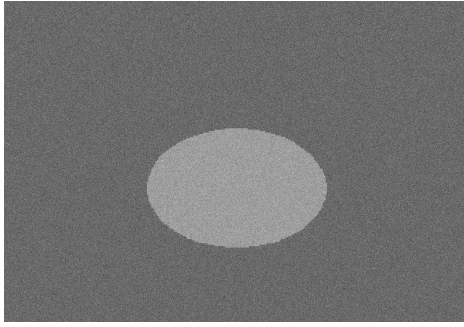
	ACC		RON		NOM		CR		DO	
	L	L1	L	L1	L	L1	L	L1	L	L1
MI	0.17	0.17	5.4	6.4	100	93	0.046	0.046	6.7	8.8
NMI	0.21	0.19	12.7	13.0	72	58	0.046	0.046	7.5	9.5
ECC	0.26	0.24	12.2	12.6	89	86	0.057	0.057	7.1	9.3
H	0.29	0.29	34.2	34.7	38	34	0.046	0.046	4.8	4.3
PSMI	0.22	0.29	34.1	34.3	50	40	0.046	0.046	1.8	1.6
E	0.29	0.29	33.2	34.0	46	37	0.046	0.046	5.9	5.2
CoR	0.14	0.14	6.2	7.4	92	84	0.046	0.057	8.8	11
WC	0.12	0	7.7	10.4	111	100	0.046	0.079	14	4.4

Table A.3: Moving image sizes, field of view (FOV), translation and rotation units of normalized parametric space and distance between two consecutive points along a line, for the five images ($N = 7$ lines, $M = 100$ sampling points and $R = 0.3$ mm).

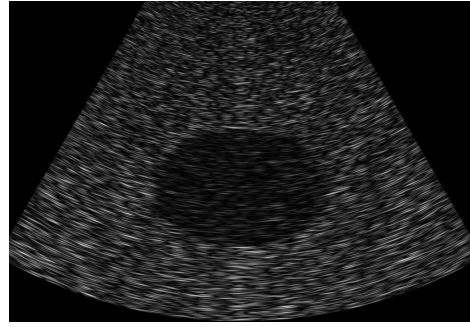
Image	FOV	Image size (mm)		Pixel size (mm)	Trans. unit (mm)	Rot. unit (rad)	δ (mm)
Oval	1	98.23	141.13	0.11	9.8	0.10694	0.0588
$LAPV_1$	Male1	78.14	112.26	0.0875	7.8	0.10729	0.0468
$LAPV_2$	Male2	78.14	112.26	0.0875	7.8	0.10729	0.0468
$LAPV_3$	Female1	74.12	106.49	0.083	7.4	0.10751	0.0444
$LAPV_4$	Female2	74.12	106.49	0.083	7.4	0.10751	0.0444

Table A.4: Values of parameters used in the simulation of CT and US images.

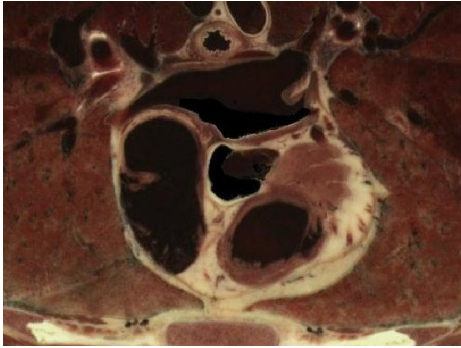
	Acoustical impedance ($\text{kg} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$)	Density ($\text{g} \cdot \text{mm}^{-3}$)	Regularity (α)	Houndsfield value
Air	4×10^{-4}	2×10^{-3}	0.1	-1000
Water	1	2×10^{-3}	0.1	0
Blood (contrast)	1.63	4×10^{-3}	0.4	400
Muscle	1.65	0.03	0.4	-20
Fat	1.35	0.1	20	-80
Bone	7.8×10^6	0.02	20	30



(a)



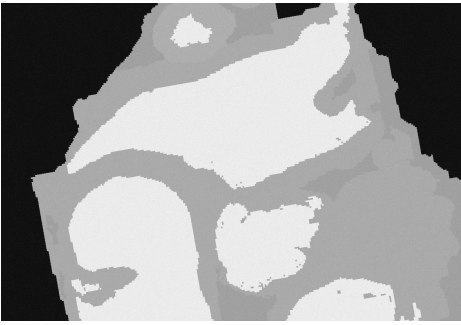
(b)



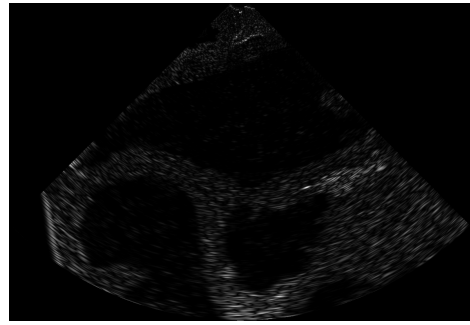
(c)



(d)



(e)



(f)

Figure A.1: Images used for evaluation. Oval geometrical form: simulated CT (a) and US (b) images. Left atrium and pulmonary veins phantom images: cryosection of male human thorax (c); segmented and labeled into six different types of tissues: air, fat, bone, muscle(including myocardium), water and blood (d); corresponding simulated CT (e) and US (f) images for a US beam depth = 7 cm.

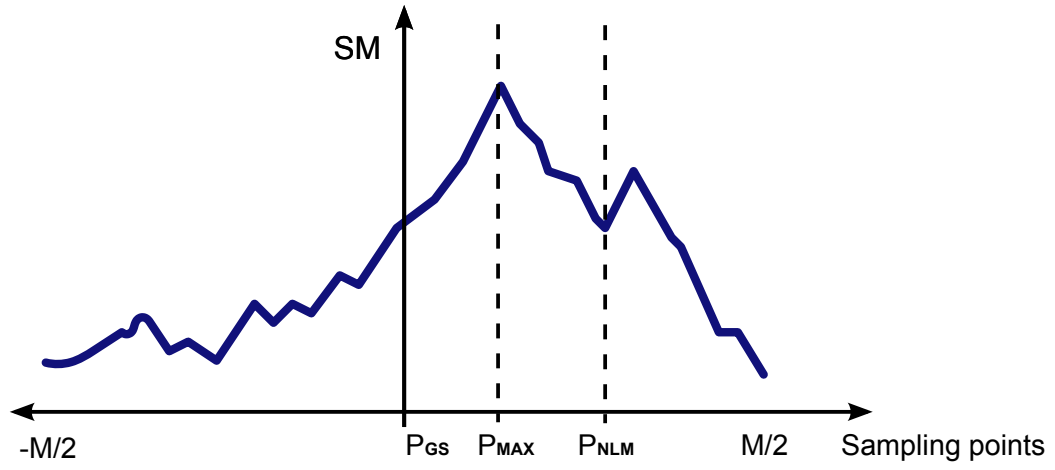


Figure A.2: Profile of a SM along one line with M sampling points, gold standard at P_{GS} , maximum value at P_{MAX} and nearest local minimum at P_{NLM} .